

The MACML Estimation of the Normally-Mixed Multinomial Logit Model

Chandra R. Bhat

The University of Texas at Austin

Department of Civil, Architectural and Environmental Engineering

1 University Station, C1761

Austin, TX 78712-0278

Tel: 512-471-4535

Fax: 512-475-8744

Email: bhat@mail.utexas.edu

January 18, 2011

ABSTRACT

The focus of this paper is to develop a procedure for the Maximum Composite Marginal Likelihood (MACML) estimation of multinomial logit models with normally mixed terms, as would be the case with normally-mixed random coefficient and/or error-component structures.

Keywords: mixed logit, multinomial logit, composite marginal likelihood, maximum simulated likelihood, discrete choice models, mixed models, panel models.

1. BACKGROUND

The Maximum Composite Marginal Likelihood (MACML) inference approach proposed by Bhat (2011) is particularly suited to estimating the multinomial probit (MNP) model, and can also be used for mixed MNP models where a normal-mixing structure (such as random coefficients) is superimposed on an MNP kernel. In the past several years, such mixed unordered-response models have mostly been estimated using a mixed (multinomial) logit formulation rather than a mixed MNP formulation, mainly because the former is easier to estimate using simulation techniques within a maximum simulated likelihood (MSL) approach. Also, in most applications of mixed models, a normally distributed mixture is used. Within the context of such normally-mixed models, the mixed MNP model is simpler to estimate using the MACML approach than is the mixed logit model (a reverse of the case with the MSL approach), because of the conjugate additional property of the normal distribution. On the other hand, the MACML estimation of the mixed logit model may be undertaken using a normal scale mixture representation for the extreme value error terms, which adds an additional layer of computational effort.

In applications, the use of a mixed MNP model or a mixed logit model is one of pure convenience, and so it stands to reason that the mixed MNP model should be used given the substantial computational efficiency achievable using the MACML inference approach compared to the MSL estimation of a mixed logit model (see Bhat and Sidharthan, 2011). However, it is of some academic interest to develop a procedure for the MACML estimation of mixed logit models. Besides, once developed, such an approach can be extended to estimating MNP models with non-normal mixing distributions. Accordingly, the focus of this paper is to develop the procedure for the MACML estimation of normally-mixed logit models.

The next section presents the MACML estimation of the mixed cross-sectional logit model, while Section 3 presents the MACML estimation of the mixed panel logit model.

2. THE MIXED CROSS-SECTIONAL MULTINOMIAL LOGIT MODEL

Consider a random-coefficients formulation in which the utility that an individual q associates with alternative i is written as:

$$U_{qi} = \beta'_q \mathbf{x}_{qi} + \varepsilon_{qi} \quad (1)$$

where \mathbf{x}_{qi} is a $(K \times 1)$ -column vector of exogenous attributes, $\boldsymbol{\beta}_q$ is an individual-specific $(K \times 1)$ -column vector of corresponding coefficients that varies across individuals based on unobserved individual attributes, and ε_{qi} is assumed to be an independently and identically distributed (across alternatives and individuals) standard Type I extreme value (or Gumbel) error term. Assume that the $\boldsymbol{\beta}_q$ vector in Equation (1) is a realization from a multivariate normal distribution with a mean vector \mathbf{b} and covariance matrix $\boldsymbol{\Omega} = \mathbf{L}\mathbf{L}'$, where \mathbf{L} is the lower-triangular Choleski factor of $\boldsymbol{\Omega}$. Then, the likelihood contribution of individual q who chooses alternative m takes the familiar mixed multinomial logit form:

$$L_q(\mathbf{b}, \boldsymbol{\Omega}) = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left(\frac{1}{1 + \sum_{i \neq m} e^{+\boldsymbol{\beta}' \mathbf{z}_{qim}}} \right) f(\boldsymbol{\beta} | \mathbf{b}, \boldsymbol{\Omega}) d\boldsymbol{\beta} = \int_{\boldsymbol{\beta}=-\infty}^{\infty} \left(\frac{e^{\boldsymbol{\beta}' \mathbf{x}_{qm}}}{\sum_i e^{\boldsymbol{\beta}' \mathbf{x}_{qi}}} \right) f(\boldsymbol{\beta} | \mathbf{b}, \boldsymbol{\Omega}) d\boldsymbol{\beta} \quad (2)$$

where $\mathbf{z}_{qim} = \mathbf{x}_{qi} - \mathbf{x}_{qm}$. This requires the integration of a K -dimensional integral, which is typically undertaken using maximum simulated likelihood estimation with pseudo-Monte Carlo (PMC) or Quasi-Monte Carlo (QMC) draws (Yu *et al.*, 2010 provides a recent overview and comparison of alternative QMC methods, while Bastin *et al.* (2006) and Qian and Shapiro (2006) discuss optimization/adaptive algorithm techniques to accelerate estimation using PMC draws). Bayesian approaches to the mixed logit estimation are also available based on Allenby and Lenk's (1994) approach (see Train, 2009, Chapter 12 for a discussion).

The MACML estimation approach we propose here (rather than the typical simulation-based approaches) for the mixed logit model is based on using a normal scale mixture representation for the extreme value error terms. Specifically, consider the utility differentials again, each between the chosen alternative m and the other $(I - 1)$ alternatives:

$$y_{qim}^* = \boldsymbol{\beta}_q' \mathbf{z}_{qim} + \varepsilon_{qi} - \varepsilon_{qm} = \mathbf{b}' \mathbf{z}_{qim} + \left(\tilde{\boldsymbol{\beta}}_q' \mathbf{z}_{qim} + \varepsilon_{qi} - \varepsilon_{qm} \right), \quad i = 1, 2, \dots, I, \quad i \neq m, \quad \boldsymbol{\beta}_q = \mathbf{b} + \tilde{\boldsymbol{\beta}}_q \quad (3)$$

The probability of alternative m being chosen continues to be given by the multivariate orthant probability expression $P_{qm} = \text{Prob}[y_{qim}^* < 0 \forall i \neq m]$. However, this orthant probability does not have a closed-form, because of the mixing of the normal term $\tilde{\boldsymbol{\beta}}_q' \mathbf{z}_{qim}$ with the extreme value terms ε_{qi} and ε_{qm} . But the approximation in Section 2.1 of Bhat (2011) based on the binary

variables holds for any multivariate cumulative distribution function (note that the derivation of the linear association coefficient vector α in Equation (6) of Bhat (2011) is independent of any distributional assumption on the vector W). Thus, we still can write the following (assuming $m \neq 1$ and $m \neq 2$ for ease in presentation and $I \geq 3$):

$$P_{qm} = \text{Prob}[y_{qim}^* < 0 \forall i \neq m] = \text{Prob}[y_{q1m}^* < 0, y_{q2m}^* < 0] \times \prod_{\substack{i=3 \\ i \neq m}}^{I-1} \left(\text{Prob}[y_{qim}^* < 0] + (\Gamma_{<i}^{-1} \Gamma_{i,<i})' \tilde{\Phi} \right), \quad (4)$$

where $\Gamma_{<i}$ and $\Gamma_{i,<i}$ are matrices of bivariate marginal probabilities defined similarly as in Section 2.1 of Bhat (2011), and $\tilde{\Phi} = \left[(1 - \text{Prob}[y_{q1m}^* < 0]), (1 - \text{Prob}[y_{q2m}^* < 0]), \dots, (1 - \text{Prob}[y_{q(I-1)m}^* < 0]) \right]'$. Again, we have only bivariate and univariate distributions in the expression. To approximate these orthant probabilities, we use the idea of a scale mixture. In particular, the bivariate probability is obtained by replacing the non-normal densities of ε_{qi} and ε_{qm} by a weighted normal mixture of component densities of the normal distribution with mean c_s and variance σ_s^2 :

$$\tilde{\lambda}(\varepsilon_{qi} = a) = \sum_{s=1}^S p_s f_N(\xi_{qi} = a; c_s, \sigma_s^2), \quad (5)$$

where $\tilde{\lambda}$ is the standard type I extreme value density function, p_s is the weight of the s^{th} mixture, and f_N is the normal density function with mean c_s and variance σ_s^2 . Equivalently,

$$\text{Prob}(\varepsilon_{qi} < a) = \sum_{s=1}^S p_s \Phi\left(\frac{a - c_s}{\sigma_s}\right).$$

Frühwirth-Schnatter and Frühwirth (2007) provide the mixture weight points p_s and the corresponding density moments (c_s and σ_s^2) based on minimizing the Kullback-Leibler distance measure (a measure of distance between the true density and estimated mixture-based density). The error in their approximation is extremely small even with 5 mixture points and literally non-existent with 10 mixture points (the values of

p_s , c_s , and σ_s^2 are tabled in their paper).¹ Now, consider a specific mixture s for the error term ε_{qi} and a specific mixture h for the error term ε_{qm} . Then for this combination of mixtures,

$$y_{qimsh}^* = \beta_q' z_{qim} + \xi_{qis} - \xi_{qmh}, \quad \beta_q \sim N(\mathbf{b}, \mathbf{\Omega}), \quad \xi_{qis} \sim N(c_s, \sigma_s^2), \quad \text{and} \quad \xi_{qmh} \sim N(c_h, \sigma_h^2) \quad (6)$$

$$\text{Thus, } P(y_{qimsh}^* < 0) = \Phi \left[\frac{-\mathbf{b}' z_{qim} - (c_s - c_h)}{\text{Var}(y_{qimsh}^*)} \right]. \quad (7)$$

To write the pairwise marginal probability $\text{Prob}(y_{qim}^* < 0, y_{qgm}^* < 0) (i \neq g)$, consider now a specific mixture l for the error term ε_{qg} , while maintaining the same mixture h for the error term ε_{qm} (this generates the needed correlation between the error terms $(\varepsilon_{qi} - \varepsilon_{qm})$ and $(\varepsilon_{qg} - \varepsilon_{qm})$). Then, we may write:

$$P(y_{qimsh}^* < 0, y_{qgmlh}^* < 0) = \Phi_2 \left[\frac{-\mathbf{b}' z_{qim} - (c_s - c_h)}{\sqrt{\text{Var}(y_{qimsh}^*)}}, \frac{-\mathbf{b}' z_{qgm} - (c_l - c_h)}{\sqrt{\text{Var}(y_{qgmlh}^*)}}, \rho_{qigmslh} \right], \quad (8)$$

$$\rho_{qigmslh} = \frac{\text{Cov}(y_{qimsh}^*, y_{qgmlh}^*)}{\sqrt{\text{Var}(y_{qimsh}^*) \text{Var}(y_{qgmlh}^*)}}, \quad (9)$$

where $\text{Var}(y_{qimsh}^*)$, $\text{Var}(y_{qgmlh}^*)$, and $\text{Cov}(y_{qimsh}^*, y_{qgmlh}^*)$ are obtained from the following (2×2) matrix:

$$\mathbf{\Sigma}_{qigslh} = \Delta_{ig} (z_q \mathbf{\Omega} z_q') \Delta_{ig}' + \mathbf{ID}_{slh}, \quad (10)$$

where Δ_{ig} is a selector matrix of size $2 \times (I-1)$ with a value of one in the i^{th} column $[(i-1)^{\text{th}}$ column] if $i < m$ $[i > m]$ ($i \neq m$), a value of one in the g^{th} column $[(g-1)^{\text{th}}$ column] if $g < m$ $[g > m]$ ($g \neq m$), and zeros everywhere else. \mathbf{ID}_{slh} is a 2×2 matrix given by:

¹ A normal mixture representation has been suggested and applied for the logistic distribution in earlier studies (Monahan and Stefanski, 1992; Drum and McCullagh, 1993; Feddag and Bacci, 2009). It is tempting to use this representation directly for the η_{qim} logistic error terms ($i = 1, 2, \dots, I, i \neq m$; $\eta_{qim} = \varepsilon_{qi} - \varepsilon_{qm}$), which we will do for the univariate distributions. However, the η_{qim} terms are correlated across the latent utility differentials because of the presence of the common extreme value ε_{qm} term. Thus, we have to derive the bivariate cumulative distribution function using a normal scale mixture for the original extreme value error terms.

$$\begin{bmatrix} \sigma_s^2 + \sigma_h^2 & \sigma_h^2 \\ \sigma_h^2 & \sigma_l^2 + \sigma_h^2 \end{bmatrix}. \quad (11)$$

Next, one can obtain the overall pairwise marginal probability by taking the weighted sum of the mixture-specific bivariate probability in Equation (8) over all permutations of h , l , and s :

$$\text{Prob}(y_{qim}^* < 0, y_{qgm}^* < 0) = \sum_{s=1}^S \sum_{l=1}^S \sum_{h=1}^S p_s p_l p_h \cdot \text{Prob}(y_{qimsh}^* < 0, y_{qgmsh}^* < 0). \quad (12)$$

The univariate marginal probabilities in Equation (4) are much simpler to approximate, since one can directly use a normal scale mixture approximation for the logistic distribution. The normal scale mixture approximation for the symmetric logistic distribution is essentially exact for as few as $\tilde{D}=5$ components (see Monahan and Stefansky, 1992; Drum and McCullagh, 1993; Feddag and Bacci, 2009). The logistic distribution may be written as follows:

$$f_L(\eta_{qim} = a) = \sum_{d=1}^{\tilde{D}} p_d f_N(\xi_{qi} = a; \sigma_d^2), \quad (13)$$

where f_L is the standard logistic density function, and f_N is the normal density function with mean a and variance σ_d^2 . Monahan and Stefansky (1992) provide the mixture weight points p_d and the corresponding density variance (σ_d^2) based on minimizing the absolute value of the difference between the true density and the estimated mixture-based density (to be precise, Monahan and Stefansky provide the mixture weight points and the inverse of the standard deviation of the corresponding density variance ($= 1/\sigma_d$)). With the above parameters, the univariate marginal probabilities may be written as follows:

$$P(y_{qim}^* < 0) = \sum_{d=1}^{\tilde{D}} p_d \Phi \left[\frac{-\mathbf{b}' \mathbf{z}_{qim}}{\text{Var}(y_{qimd}^*)} \right], \text{ where } \text{Var}(y_{qimd}^*) = \mathbf{\Delta}_i (\mathbf{z}_q \mathbf{\Omega} \mathbf{z}_q') \mathbf{\Delta}_i' + \sigma_d^2, \quad (14)$$

and $\mathbf{\Delta}_i$ is a selector matrix of size $1 \times (I-1)$ with a value of one in the i^{th} column $[(i-1)^{\text{th}}$ column] if $i < m$ [$i > m$] ($i \neq m$), and zeros everywhere else.

With the expressions above, the likelihood of individual q in the mixed multinomial logit model is obtained from Equation (4).

3. THE MIXED PANEL MULTINOMIAL LOGIT MODEL

Consider the following model with ‘ t ’ now being an index for choice occasion:

$$U_{qit} = \beta'_q \mathbf{x}_{qit} + \varepsilon_{qit}, \quad \beta_q \sim MVN(\mathbf{b}, \mathbf{\Omega}), \quad q = 1, 2, \dots, Q, \quad i = 1, 2, \dots, I, \quad t = 1, 2, \dots, T \quad (15)$$

Let ε_{qit} be IID standard Type I-extreme value distributed over individuals, alternatives, and choice occasions. We will assume that the coefficients β_q are constant over choice situations of a given decision maker. Also, we will assume that the number of choice occasions per individual is the same across all individuals. The case of different numbers of choice occasions per individual may be handled in a straightforward fashion based on Section 4.2 of Bhat (2011).

The likelihood contribution of individual q who selects alternative m_t at the t^{th} choice occasion t ($t = 1, 2, \dots, T$) is:

$$L_q(\mathbf{b}, \mathbf{\Omega}) = \int_{\beta=-\infty}^{\infty} \prod_{t=1}^T \left(\frac{1}{1 + \sum_{i \neq m} e^{\beta' \mathbf{z}_{qim_t}}} \right) f(\beta | \mathbf{b}, \mathbf{\Omega}) d\beta = \int_{\beta=-\infty}^{\infty} \prod_{t=1}^T \left(\frac{e^{\beta' \mathbf{x}_{qm_t}}}{\sum_i e^{\beta' \mathbf{x}_{qit}}} \right) f(\beta | \mathbf{b}, \mathbf{\Omega}) d\beta, \quad (16)$$

where $\mathbf{z}_{qim_t} = (\mathbf{x}_{qit} - \mathbf{x}_{qm_t})$. For the estimation of the above mixed panel MNL model, we bring together all the three devices of (1) approximation for multivariate orthant probabilities, (2) the CML approach, and (3) the scale-mixture technique. In the mixed cross-sectional MNL mode, we started from the normal scale mixture approximation for the type I extreme value error term, because (in the notation of the panel model now) the utility difference error terms η_{qim_t} at time t ($\eta_{qim_t} = \varepsilon_{qit} - \varepsilon_{qm_t}; i = 1, 2, \dots, I, i \neq m_t$) are correlated by the presence of the common error term ε_{qm_t} . This is still the case for the bivariate probabilities involving the latent utility differentials within the same time period. However, there is an important way to increase computational efficiency in the CML estimation of the mixed panel MNL model when considering pairwise marginal likelihoods across time periods. In particular, note that the utility difference error terms η_{qim_t} ($i \neq m_t$) and η_{qgm_w} ($g \neq m_w$), $t \neq w$, are not correlated across time periods. This recognition can be gainfully employed, so that when computing the bivariate marginal likelihoods in the approximation method for the inter-temporal terms in the MACML approach, we can directly use a normal scale mixture approximation for the logistic distribution.

Using the same notations and derivational approach as for the mixed cross-sectional MNL model, the relevant bivariate marginal orthant probabilities for the latent utilities within the same time period t take the form below ($i, g \neq m_t, i \neq g$):

$$\begin{aligned} & \text{Prob}[y_{qim_t}^* < 0, y_{qgm_t} < 0] \\ &= \sum_{s=1}^S \sum_{l=1}^S \sum_{h=1}^S p_s p_l p_h \Phi_2 \left[\frac{-\mathbf{b}' \mathbf{z}_{qim_t} - (c_s - c_h)}{\sqrt{\text{Var}(y_{qim_tsh}^*)}}, \frac{-\mathbf{b}' \mathbf{z}_{qgm_t} - (c_l - c_h)}{\sqrt{\text{Var}(y_{qgm_tlh}^*)}}, \frac{\text{Cov}(y_{qim_tsh}^*, y_{qgm_tlh}^*)}{\sqrt{\text{Var}(y_{qim_tsh}^*) \text{Var}(y_{qgm_tlh}^*)}} \right] \end{aligned} \quad (17)$$

$\text{Var}(y_{qim_tsh}^*)$, $\text{Var}(y_{qgm_tlh}^*)$, and $\text{Cov}(y_{qim_tsh}^*, y_{qgm_tlh}^*)$ in the expression below are obtained from the following (2×2) matrix:

$$\Sigma_{qigtsh} = \Delta_{igt} (\tilde{\mathbf{z}}_q \mathbf{\Omega} \tilde{\mathbf{z}}_q') \Delta_{igt}' + \begin{bmatrix} \sigma_s^2 + \sigma_h^2 & \sigma_h^2 \\ \sigma_h^2 & \sigma_l^2 + \sigma_h^2 \end{bmatrix}, \quad (18)$$

where $\tilde{\mathbf{z}}_q$ is a $[T \times (I-1)] \times K$ matrix obtained by vertically concatenating the transpose of the $K \times 1$ vectors \mathbf{z}_{qim_t} ($i = 1, 2, \dots, I, i \neq m_t; t = 1, 2, \dots, T$) (note that there are $T \times (I-1)$ vectors in \mathbf{z}_{qim_t}), and Δ_{igt} is a selection matrix of size $2 \times [(I-1) \times T]$. It has a value of one in the $[(I-1) \times (t-1) + i]^{\text{th}}$ column in the first row if $i < m_t$ or a value of one in the $[(I-1) \times (t-1) + (i-1)]^{\text{th}}$ column in the first row if $i > m_t$. Similarly, it has a value of one in the $[(I-1) \times (t-1) + g]^{\text{th}}$ column in the second row if $i < m_t$ or a value of one in the $[(I-1) \times (t-1) + (g-1)]^{\text{th}}$ column in the second row if $i > m_t$. The matrix has values of zero everywhere else.

The corresponding bivariate orthant probability expression for the pairings across time periods, using the normal scale mixture representation directly for the independent logistic terms is (for $t \neq w: i \neq m_t$ and $g \neq m_w$):

$$\begin{aligned} & \text{Prob}[y_{qim_t}^* < 0, y_{qgm_w} < 0] \\ &= \sum_{d=1}^D \sum_{e=1}^D p_d p_e \Phi_2 \left[\frac{-\mathbf{b}' \mathbf{z}_{qim_t} - c_d}{\sqrt{\text{Var}(y_{qim_td}^*)}}, \frac{-\mathbf{b}' \mathbf{z}_{qgm_w} - c_e}{\sqrt{\text{Var}(y_{qgm_w e}^*)}}, \frac{\text{Cov}(y_{qim_td}^*, y_{qgm_w e}^*)}{\sqrt{\text{Var}(y_{qim_td}^*) \text{Var}(y_{qgm_w e}^*)}} \right] \end{aligned} \quad (19)$$

$\text{Var}(y_{qim,td}^*)$, $\text{Var}(y_{qgm_w,we}^*)$, and $\text{Cov}(y_{qim,td}^*, y_{qgm_w,we}^*)$ are obtained from the following (2×2) matrix:

$$\Sigma_{qigtwde} = \Delta_{igtw} (\tilde{z}_q \Omega \tilde{z}_q') \Delta_{igtw}' + \begin{bmatrix} \sigma_d^2 & 0 \\ 0 & \sigma_e^2 \end{bmatrix}, \quad (20)$$

In the above expression, Δ_{igtw} is a selection matrix of size $2 \times [(I-1) \times T]$. This matrix has a value of one in the $[(I-1) \times (t-1) + i]^{\text{th}}$ column in the first row if $i < m_t$ or a value of one in the $[(I-1) \times (t-1) + (i-1)]^{\text{th}}$ column in the first row if $i > m_t$. Similarly, it has a value of one in the $[(I-1) \times (w-1) + g]^{\text{th}}$ column in the second row if $g < m_w$ or a value of one in the $[(I-1) \times (w-1) + (g-1)]^{\text{th}}$ column in the second row if $g > m_w$. The matrix has values of zero everywhere else.

4. CONCLUSIONS

In this paper, we introduce a simulation-free maximum approximated composite marginal likelihood (MACML) estimation approach for normally-mixed multinomial logit models in both a cross-sectional and panel context. The approach is based on using a normal scale mixture representation for the extreme value error terms. In both the cross-sectional and panel contexts, the MACML approach involves only univariate and bivariate cumulative normal distribution function evaluations, regardless of the number of alternatives or the number of choice occasions per individual. Extensions to include more complex mixing in panel models, such as intra- and cross-temporal random coefficients and/or autoregressive random coefficients, pose no conceptual difficulties and can be based on the same normal scale mixture technique as in the simpler mixing cases. The scale mixture technique developed here can also be extended to handle non-normal mixing distributions with an MNP kernel or a logit kernel.

REFERENCES

- Allenby, G.M., Lenk, P.J., 1994. Modeling household purchase behavior with logistic normal regression. *Journal of the American Statistical Association* 89(428), 1218-1231.
- Bastin, F., Cirillo, C., Toint, P.L., 2006. Application of an adaptive Monte Carlo algorithm to mixed logit estimation. *Transportation Research Part B* 40(7), 577-593.
- Bhat, C.R., 2011. The maximum approximated composite marginal likelihood (MACML) estimation of multinomial probit-based unordered response choice models. Technical Paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin.
http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/MACML_unordered_response_paper.pdf.
- Bhat, C.R., Sidharthan, R., 2011. A simulation evaluation of the maximum approximated composite marginal likelihood (MACML) estimator for mixed multinomial probit models. Technical paper, Department of Civil, Architectural & Environmental Engineering, The University of Texas at Austin.
http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/mixedmodelcml_simulation.pdf.
- Drum, M.L., McCullagh, P., 1993. REML estimation with exact covariance in the logistic mixed model. *Biometrics* 49(3), 677-689.
- Feddag, M.-L., Bacci S., 2009 Pairwise likelihood for the longitudinal mixed Rasch model, *Computational Statistics and Data Analysis* 53(4), 1027-1037.
- Frühwirth-Schnatter, S., Frühwirth, R., 2007. Auxiliary mixture sampling with applications to logistic models. *Computational Statistics and Data Analysis* 51(7), 3509-3528.
- Monahan, J.F., Stefanski, L.A., 1992. Normal scale mixture approximations to $F^*(x)$ and computation of the logistic-normal integral. In *Handbook of the Logistic Distribution*, Balakrishnan, N. (ed), Marcel Dekker, Inc., New York, 529-540.
- Qian, Z., Shapiro, A., 2006. Simulation-based approach to estimation of latent variable models. *Computational Statistics and Data Analysis* 51(2), 1243-1259.
- Train, K. 2009. *Discrete Choice Methods with Simulation*, 2nd ed., Cambridge University Press, Cambridge.
- Yu, J., Goos, P., Vandebroek, M., 2010. Comparing different sampling schemes for approximating the integrals involved in the efficient design of stated choice experiments. *Transportation Research Part B* 44(10), 1268-1289.